

Перевод на английский язык <https://vavilov.elpub.ru/jour>

InterTransViewer: сравнительное описание профилей дифференциальной экспрессии генов из разных экспериментов

А.В. Тяпкин^{1, 2}, В.В. Лавреха^{1, 2}, Е.В. Убогоева¹, Д.Ю. Ощепков¹, Н.А. Омелянчук¹, Е.В. Землянская^{1, 2} 

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

 ezemlyanskaya@bionet.nsc.ru

Аннотация. В настоящее время в связи со стремительным ростом количества полногеномных экспериментов по изучению изменения экспрессии генов в различных условиях все более широкое распространение получают методы метаанализа транскриптомных данных из разных экспериментов, так как интеграция данных может обеспечить большую точность в выявлении генов-кандидатов и позволяет тестировать новые гипотезы, которые невозможно было проверить в отдельных исследованиях. Для повышения информативности такой интеграции необходимо оптимизировать подбор экспериментов. В настоящей работе мы предлагаем набор количественных показателей для всестороннего сравнительного описания транскриптомных данных. Эти показатели легко могут быть визуализированы и интерпретированы. Они включают в себя количество дифференциально экспрессирующихся генов (ДЭГ), долю специфических (уникальных) ДЭГ в каждом наборе данных, попарное сходство экспериментов по составу ДЭГ, оценку однородности профилей дифференциально экспрессирующихся генов. Для автоматического вычисления и визуализации этих показателей мы разработали программу InterTransViewer. Мы применили InterTransViewer для сравнительного описания транскрипционных ответов на обработку фитогормонами у модельного растения *Arabidopsis thaliana* L., взяв в анализ 23 единообразно обработанных профиля дифференциальной экспрессии генов в ответ на ауксин и 16 профилей дифференциальной экспрессии, индуцированных этиленом или его предшественником – 1-аминоциклопропановой кислотой. Мы продемонстрировали, что комплексное рассмотрение характеристик отдельных профилей ДЭГ в контексте результатов попарных сравнений профилей по составу ДЭГ позволяет позиционировать эксперименты в контексте друг друга, оценивать тенденцию к их интеграции или сегрегации, генерировать гипотезы о влиянии весомых нецелевых факторов на исследуемый транскрипционный ответ. В результате это дает возможность выделять потенциально однородные группы экспериментов. Последующий анализ однородности этих групп профилей с помощью процедуры ресемплинга и установления порога уровня значимости помогает принять решение о целесообразности использования этих данных для метаанализа. В целом InterTransViewer позволяет эффективно формировать выборки экспериментов в зависимости от задачи и методов метаанализа. Ключевые слова: транскриптом; интеграция данных; ауксин; этилен; *Arabidopsis thaliana* L.


Для цитирования: Тяпкин А.В., Лавреха В.В., Убогоева Е.В., Ощепков Д.Ю., Омелянчук Н.А., Землянская Е.В. InterTransViewer: сравнительное описание профилей дифференциальной экспрессии генов из разных экспериментов. *Вавиловский журнал генетики и селекции*. 2023;27(8):1042-1052. DOI 10.18699/VJGB-23-119

InterTransViewer: a comparative description of differential gene expression profiles from different experiments

А.В. Тяпкин^{1, 2}, В.В. Лавреха^{1, 2}, Е.В. Убогоева¹, Д.Ю. Ощепков¹, Н.А. Омелянчук¹, Е.В. Землянская^{1, 2} 

¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

 ezemlyanskaya@bionet.nsc.ru

Abstract. Meta-analysis of transcriptomic data from different experiments has become increasingly prevalent due to a significantly increasing number of genome-wide experiments investigating gene expression changes under various conditions. Such data integration provides greater accuracy in identifying candidate genes and allows testing new hypotheses, which could not be validated in individual studies. To increase the relevance of experiment integration, it is necessary to optimize the selection of experiments. In this paper, we propose a set of quantitative indicators for a comprehensive comparative description of transcriptomic data. These indicators can be easily visualized and interpreted. They include the number of differentially expressed genes (DEGs), the proportion of experiment-specific (unique) DEGs in each data set, the pairwise similarity of experiments in DEG composition and the homogeneity of DEG profiles. For automatic calculation and visualization of these indicators, we have developed the program InterTransViewer. We have used InterTransViewer to comparatively describe 23 auxin- and 16 ethylene- or 1-aminocyclopropane-1-carboxy-

lic acid (ACC)-induced transcriptomes in *Arabidopsis thaliana* L. We have demonstrated that analysis of the characteristics of individual DEG profiles and their pairwise comparisons based on DEG composition allow the user to rank experiments in the context of each other, assess the tendency towards their integration or segregation, and generate hypotheses about the influence of non-target factors on the transcriptional response. As a result, InterTransViewer identifies potentially homogeneous groups of experiments. Subsequent estimation of the profile homogeneity within these groups using resampling and setting a significance threshold helps to decide whether these data are appropriate for meta-analysis. Overall, InterTransViewer makes it possible to efficiently select experiments for meta-analysis depending on its task and methods.

Key words: transcriptome; data integration; auxin; ethylene; *Arabidopsis thaliana* L.

For citation: Тяпкин А.В., Лавреха В.В., Убогоева Е.В., Ощепков Д.Ю., Омелянчук Н.А., Землянская Е.В. InterTransViewer: a comparative description of differential gene expression profiles from different experiments. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2023;27(8):1042-1052. DOI 10.18699/VJGB-23-119

Введение

На сегодняшний день анализ дифференциальной экспрессии генов в различных условиях – это один из важнейших подходов к исследованию генетической регуляции формирования признаков (Stelpflug et al., 2016; Tello-Ruiz et al., 2016). Стремительный рост количества экспериментов по полногеномному профилированию экспрессии генов в различных условиях и доступность их результатов в базах данных по функциональной геномике, таких как Gene Expression Omnibus (GEO) (Clough, Barrett, 2016) или BioStudies (Sarkans et al., 2021), открывают широкий простор для сравнительного анализа результатов экспериментов из различных исследований с целью их обобщения с помощью методов метаанализа (Cahan et al., 2007; Rung, Brazma, 2013; Keel, Lindholm-Perry, 2022). Такой подход позволяет не только определять наиболее надежные дифференциально экспрессирующиеся гены (ДЭГ), стабильно определяющиеся в независимых исследованиях (Freire-Rios et al., 2020), но и увеличивать выборку образцов с целью выявления слабых закономерностей (Bairakdar et al., 2023) или тестирования гипотез, которые невозможно было исследовать в одиночных работах (Sudmant et al., 2015; Winter et al., 2019).

Для успешной интеграции данные должны удовлетворять определенным критериям (Cahan et al., 2007; Rung, Brazma, 2013; Yu, Zeng, 2018). В первую очередь, они должны быть исчерпывающе описаны в соответствии с установленными минимальными требованиями к описанию транскриптомных экспериментов (Brazma et al., 2001; Brazma, 2009). Кроме того, объединяемые эксперименты должны исследовать аналогичные гипотезы о влиянии одного и того же фактора. При этом следует избегать или корректировать так называемый пакетный эффект (batch effect), когда нецелевые факторы (биологические характеристики объекта, условия проведения эксперимента, протокол пробоподготовки, выбор платформы для получения данных и пр.) влияют на результаты конкретного эксперимента.

Простая фильтрация данных по экспериментальным условиям не всегда обеспечивает оптимальный выбор данных для метаанализа. С одной стороны, весомый нецелевой фактор может отсутствовать в описании данных, в результате формальное совпадение условий экспериментов не исключает пакетного эффекта. С другой стороны, результаты экспериментов, выполненных в неидентичных условиях, могут быть достаточно хорошо согласованы. Сравнительное описание транскриптомных данных из

разных экспериментов позволяет оптимизировать выбор данных и методов для их предобработки. Однако до сих пор для этой процедуры не разработан стандарт, а также существенно ощущается нехватка программных средств, в особенности для графического представления результатов. Например, программа MetaQC, применяемая для оценки качества микрочипов, оценивает шесть количественных метрик: (1) воспроизводимость коэкспрессирующихся групп генов в различных экспериментах; (2) согласованность паттерна коэкспрессии известных генов с базами данных метаболических и сигнальных путей (т. е. участие генов в одном процессе); (3–4) точность выявления обогащения группы ДЭГ в терминах генной онтологии (т. е. участие генов в процессах, связь с клеточными компонентами или молекулярными функциями) и их согласованность в разных экспериментах; 5) правильность определения известных биомаркеров; 6) согласованность ранжирования ДЭГ между транскриптомами (Kang et al., 2012). Однако MetaQC не позволяет визуализировать эти метрики, и некоторые из метрик качества опираются на внешние базы данных и известные маркеры, а не на внутренние характеристики профилей экспрессии, что может затенять малоизученные процессы и усложнять анализ для немодельных видов.

Другая программа, ViDGER, созданная для упрощения интерпретации данных экспериментов по секвенированию РНК, предоставляет широкий спектр визуализаций, но не предлагает удобных средств для сравнения профилей ДЭГ (McDermaid et al., 2019). В программе NetworkAnalyst 3.0 акцент сделан на реконструкцию сетей межбелковых взаимодействий, но она также предусматривает некоторую возможность визуального сравнения списков генов с помощью интерактивных тепловых карт, сетей обогащения, диаграмм Венна и хордовых диаграмм (Zhou et al., 2019).

В настоящей работе мы предлагаем набор показателей для всестороннего сравнительного описания профилей ДЭГ, которые легко могут быть визуализированы и интерпретированы. Они характеризуют отдельные профили дифференциальной экспрессии, их попарное сходство и тенденцию к интеграции или сегрегации. Для автоматического вычисления и визуализации этих показателей мы разработали программу InterTransViewer, которую применили для сравнительного описания транскрипционных ответов на обработку фитогормонами ауксином (23 профиля ДЭГ из 16 исследований) и этиленом (16 профилей ДЭГ из восьми исследований) у модельного вида растений *Arabidopsis thaliana* L.

Материалы и методы

Характеристики отдельных профилей дифференциальной экспрессии. Для каждого индивидуального списка ДЭГ мы оценивали: 1) количество ДЭГ, 2) долю ДЭГ, специфичных для данного списка, а также 3) отношение доли специфических ДЭГ к доле ДЭГ данного списка среди генов, идентифицированных как ДЭГ хотя бы в одном из анализируемых списков:

$$R_i = \frac{\delta_i \cdot N}{n_i},$$

где R_i – отношение долей для списка i ; δ_i – доля ДЭГ, специфичных для списка i ; n_i – количество ДЭГ в списке i ; N – количество ДЭГ в общем списке. Вычисленные показатели графически представляются в виде зеркальных гистограмм. Они в совокупности с метаданными экспериментов позволяют в первом приближении оценить сходство списков и определить потенциальные выбросы. Например, слишком малое/большое количество ДЭГ или высокое значение R , которое не коррелирует со специфическими условиями эксперимента или биологическими свойствами объекта, может свидетельствовать о влиянии неизвестного весомого нецелевого фактора или о низком качестве данных.

Попарные сравнения профилей дифференциальной экспрессии по составу ДЭГ. Если список ДЭГ меньшего размера вложен в другой список ДЭГ большего размера, и отклонение размера каждого списка от среднего незначительно или коррелирует со специфическими условиями эксперимента или биологическими свойствами объекта, мы считаем, что результаты двух экспериментов консистентны. Чтобы оценить сходство двух профилей с учетом этого условия, для каждой пары списков ДЭГ мы рассчитали индекс сходства I следующим образом:

$$I = \frac{c}{\min\{a, b\} + c},$$

где c – это количество общих ДЭГ между списками; a – количество ДЭГ, представленных в первом списке и отсутствующих во втором; b – количество ДЭГ, представленных во втором списке и отсутствующих в первом. Таким образом, индекс сходства I отражает долю общих ДЭГ в меньшем списке. Индекс сходства может принимать значения от нуля до единицы, при этом ноль соответствует отсутствию общих ДЭГ в двух списках, а единица – полному вложению одного списка в другой. Матрицы сходства списков ДЭГ визуализируются в виде тепловой карты, на основании которой можно не только сделать вывод о сходстве профилей экспрессии по составу ДЭГ, но и выделить отдельные группы наиболее сходных экспериментов.

Кластеризация профилей дифференциальной экспрессии. Матрица сходства, описанная выше, сопоставляет списки ДЭГ, не учитывая величину изменения уровня экспрессии генов. Чтобы выявлять группы сходных профилей дифференциальной экспрессии с учетом этой величины, мы использовали иерархическую кластеризацию на основании матрицы евклидовых расстояний в пространстве \log_2 -трансформированных значений изменения уровней экспрессии генов, идентифицируемых как ДЭГ хотя бы в одном из анализируемых списков, без учета значимости этих изменений. Предварительно эти значе-

ния были нормированы на диапазон разброса в каждом эксперименте и стандартизованы для каждого гена, что позволило сравнивать профили транскрипционного ответа из разных экспериментов. Иерархическую кластеризацию осуществляли с помощью функции `Vclust` из пакета `shirunov v.1.17.1` (<https://CRAN.R-project.org/package=shirunov>), применяя метод `Ward.D2`, основанный на минимизации суммы квадратов евклидовых расстояний между каждым объектом кластера и центроидом кластера.

Количественная оценка однородности группы профилей по составу ДЭГ. Пусть A – множество генов, идентифицируемых как ДЭГ хотя бы в одном из m анализируемых списков, а количество этих генов $|A| = N$. Множество A включает в себя: 1) ДЭГ, изменение уровня экспрессии которых в заданной выборке из m списков определяется преимущественно действием целевого фактора, а также 2) гены, изменение уровня экспрессии которых существенно зависит от нецелевых факторов. Очевидно, что, если вычислить значение N_k для подвыборки из k списков ($k < m$) и далее вычислять значения N_{k+i} , добавляя к этой подвыборке по одному списку, то величина N_{k+i} должна не убывать с ростом i . При этом чем более неоднороден исследуемый набор профилей ДЭГ (чем больше он содержит списков, сформированных под влиянием различающихся нецелевых факторов), тем сильнее будет рост величины N_{k+i} .

Используя ресемплинг, мы создали $m - 1$ наборов псевдовыборок списков ДЭГ: в одном наборе i ($i \in \mathbb{N}$, $i = [1; m - 1]$) каждая псевдовыборка состояла из $k_i < m$ списков ДЭГ ($k_1 = m - 1$, $k_{i+1} = k_i - 1$), чтобы оценить, при каком значении k_i будет наблюдаться значимое снижение N_{k_i} по сравнению с N_m . Для формирования одной псевдовыборки из оригинального набора списков ДЭГ, состоящего из m элементов, мы случайным образом выбирали с возвращением k_i списков (рис. 1), для которых далее определяли количество генов $N_{k_{ij}}$, идентифицируемых как ДЭГ хотя бы в одном из k_i списков (индекс j обозначает номер псевдовыборки в одном наборе). Одновременно мы создавали псевдовыборку из m списков и рассчитывали соответствующее значение N_{mj} , после чего вычисляли разность $d_j = N_{mj} - N_{k_{ij}}$ (см. рис. 1). В результате 5000 итераций ($j \in \mathbb{N}$, $j = [1; 5000]$) формировался вариационный ряд этих разностей. Доверительный интервал определяли методом процентилей (Rousset et al., 2021). Если при каких-то значениях k_i наблюдалась значимая разница между N_m и N_{k_i} , анализируемый набор профилей считали неоднородным. Распределение значений d визуализировали в виде гистограммы.

Программная реализация InterTransViewer. Программа `InterTransViewer` реализована в виде скрипта на языке R (v.4.1.2) и доступна по ссылке (<https://github.com/al-t1/InterTransViewer0/>). `InterTransViewer` принимает на вход таблицу, в которой первый столбец содержит одну группирующую переменную (идентификаторы генов, ID), а каждая следующая пара столбцов содержит \log_2 -трансформированные значения изменений уровней экспрессии генов (\log_2FC) и соответствующие им скорректированные значения p (p -adjusted) для отдельного эксперимента. Если пользователь осуществляет предобработку сырых транскриптомных данных самостоятельно, такая таблица может

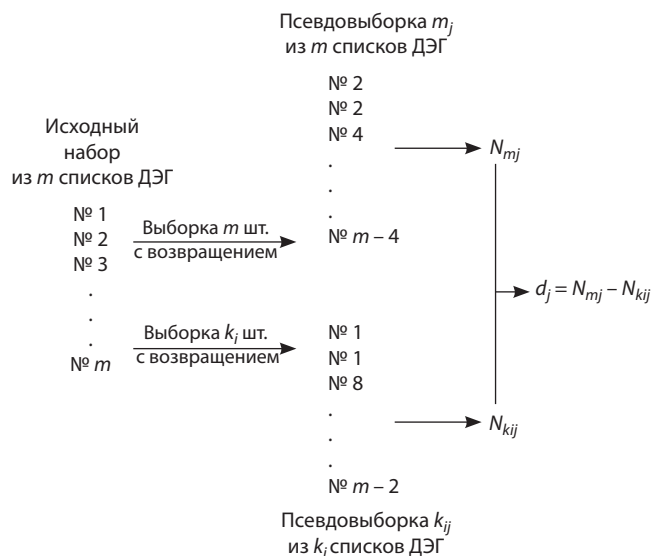


Рис. 1. Процедура создания двух псевдовыборок, каждая из которых состоит из m и k_j ($k_j < m$) случайно выбранных списков ДЭГ с возвращением, и определения разницы d_j между количеством генов, идентифицируемых как ДЭГ хотя бы в одном из m и k_j списков (N_{mj} и N_{kij} соответственно).

Индекс j обозначает номер псевдовыборки. Операцию повторяли 5000 раз ($j \in \mathbb{N}$, $j = [1; 5000]$), генерируя таким образом распределение значений d_j , что позволяло оценить значимость различия значений N в псевдовыборках. Такая процедура повторялась для каждого значения k_j ($k_1 = m - 1$; $k_{i+1} = k_i - 1$, где $i \in \mathbb{N}$, $i = [1; m - 1]$).

быть собрана с использованием функции InterTrans Viewer DEGweave для объединения результатов, сгенерированных функцией topTable limma и/или функцией results пакета DESeq2 в процессе предобработки сырых данных микрочипов и РНК-секвенирования соответственно. Желательно, чтобы предварительная обработка сырых данных была выполнена максимально единообразно для каждой технологической платформы, а также чтобы дизайн каждого эксперимента предусматривал как минимум по две биологические реплики для контрольных и опытных образцов.

На этапе контроля качества данных рекомендуется обращать особое внимание на разброс данных для реплик: например, для микрочипов с использованием функции plotMDS пакета limma (Ritchie et al., 2015); с помощью программ fastQC и fastp для сырых данных секвенирования РНК (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>; Chen et al., 2018) и с использованием функции plotPCA пакета DESeq2 (Love et al., 2014) для матрицы количества прочтений, соответствующих гену. Необходимо, чтобы все профили дифференциальной экспрессии отражали действие одного целевого фактора. Формально профиль дифференциальной экспрессии подойдет для анализа с помощью InterTransViewer, если в нем есть хотя бы один ДЭГ для установленного уровня значимости, однако рекомендуется более жестко ограничивать этот параметр, выбирая в качестве порогового значения не менее 10.

Вычисление показателей для сравнительного описания профилей ДЭГ и их визуализация реализованы в

виде функций, описанных в документации программы InterTransViewer. Например, количество ДЭГ, долю эксперимент-специфичных ДЭГ и отношение R_i для всех экспериментов можно получить с помощью функции DEGsummary, после чего представить результаты в виде столбчатых диаграмм с использованием функций TotalSpecPlot и RmetricPlot. Функция GetSimMatrix позволяет получить матрицу сходства I . Функция DE_bootstrap дает возможность проводить ресемплинг по описанной выше методике. Иерархическая кластеризация выполняется с помощью функции DE_clustering.

В результате работы программа генерирует широкий спектр выходных данных. Для каждого эксперимента формируются две таблицы, содержащие список ДЭГ (для заданного порогового значения p -adj и, если необходимо, logFC) и список эксперимент-специфичных ДЭГ с указанием logFC и p -adj. Также сохраняются общий список генов, которые являются ДЭГ хотя бы в одном эксперименте с указанием количества экспериментов, в которых ген определен как ДЭГ; сводная таблица, сгенерированная функцией DEGsummary, и соответствующие гистограммы; матрица сходства I и тепловая карта; дендрограммы, полученные при кластеризации; таблицы и диаграммы с результатами ресемплинга для оценки однородности наборов профилей дифференциальной экспрессии.

Транскриптомные данные из открытых источников.

Мы собрали все доступные в открытых базах транскриптомные данные по обработке *A. thaliana* фитогормонами ауксином, этиленом, их предшественниками или синтетическими аналогами. Из них мы отобрали транскриптомные целых проростков или отдельных органов растений дикого типа, в которых обработки гормонами были дополнены контрольными экспериментами (псевдообработка или отсутствие обработки). Чтобы сделать возможным последующий сравнительный анализ, мы осуществили единообразную предобработку сырых данных. Данные микрочип-экспериментов были загружены из базы данных GEO (<https://www.ncbi.nlm.nih.gov/geo/>). Данные секвенирования РНК были взяты из базы данных NCBI Sequence Read Archive (SRA) (<https://www.ncbi.nlm.nih.gov/sra/>). Последовательность генома *A. thaliana* и его аннотация (TAIR 10) были загружены из Ensembl Plants (<https://plants.ensembl.org/index.html>, release 52).

Все собранные микрочип-эксперименты выполнены на платформе ATH1. Нормировка сырых данных микрочипов и определение ДЭГ осуществлялись с помощью пакета limma v.3.52.4 (Ritchie et al., 2015). Для контроля качества данных секвенирования РНК использована программа FastQC v.0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Прочтения, полученные на платформе Illumina, подвергались триммингу и фильтрации по качеству с помощью программы fastp v.0.23.2 (Chen et al., 2018), запускаемой со следующими параметрами: -q 20 -u 30 -5 -3 -W 4 -M 20, после чего выполнялось выравнивание прочтений на геном *A. thaliana* с применением HISAT2 v.2.2.1 (Kim et al., 2019). Прочтения SOLiD выравнивались на геном с помощью программы TopHat (Kim et al., 2013). Для подсчета числа уникально картированных прочтений использовали функцию summarizeOverlaps из пакета GenomicAlignments R v.1.30.0

Данные микрочип-экспериментов и РНК-секвенирования, использованные в исследовании

№ эксперимента	Название эксперимента	Тип	Ткань, стадия развития	Обработка (концентрация, время)	Число реплик	Литературный источник
Ауксин						
1	ERP021928	P	Меристема и молодые цветы до и во время 10-й стадии	10 мкМ ИУК, 0,5 ч	3	Simonini et al., 2017
2, 4, 9	GSE18975	M	7-дневные проростки	1 мкМ ИУК, 0,5 ч; 1 мкМ ИУК, 1 ч; 1 мкМ ИУК, 3 ч	3	Delker et al., 2010
3	SRP258689	P	Корни 3-дневных проростков	1 мкМ ИУК, 1 ч	3	Freire-Rios et al., 2020
5, 15	GSE3350	M	Корни без корневого апекса 3-дневных проростков, выращенных на среде MS с 10 мкМ НФК	10 мкМ ИУК, 2 ч; 10 мкМ ИУК, 6 ч	2	Vanneste et al., 2005
6	GSE35580	M	Корни 7-дневных проростков	5 мкМ ИУК, 2 ч	3	Bargmann et al., 2013
7, 16	GSE42896	M	Корни без корневого апекса 3-дневных проростков, выращенных на среде MS с 10 мкМ НФК	10 мкМ ИУК, 2 ч; 10 мкМ ИУК, 6 ч	3	De Rybel et al., 2012
8	GSE627	M	7-дневные проростки	5 мкМ ИУК, 2 ч	3	Okushima et al., 2005
10	GSE58028	M	7- и 8-дневные проростки	1 мкМ ИУК, 3 ч	3	
11	SRP033494	P	Корни 7-дневных проростков	5 мкМ ИУК, 4 ч	2	Chaiwanon, Wang, 2015
12, 18, 19, 21	GSE42007	M	Корни 6-дневных проростков	1 мкМ ИУК, 4 ч; 1 мкМ ИУК, 8 ч; 1 мкМ ИУК, 12 ч; 1 мкМ ИУК, 24 ч	3	Lewis et al., 2013
13	GSE7432	M	Корни 3-дневных этилированных проростков	1 мкМ ИУК, 4 ч	2	Stepanova et al., 2007
14	SRP102803	P	Корни 3-дневных проростков	1 мкМ ИУК, 6 ч	3	Omelyanchuk et al., 2017
17	GSE59426	M	Кончики корней 3-дневных проростков	10 мкМ ИИМК, 6 ч	3	Xuan et al., 2015
20	GSE59741	M	Пазушные почки листьев 21- и 28-дневных проростков	1 мкМ ИУК, 18 ч	3	Müller et al., 2015
22	SRP074436	P	Апексы побегов и пазушные меристемы 14-дневных проростков	5 мкМ 2,4-Д, 55 ч	3	Mozgová et al., 2017
23	GSE179303	M	Листья одинакового размера и стадии развития	23 мМ 2,4-Д, 72 ч после обработки (опрыскивание)	3	Romero-Puertas et al., 2022
Этилен						
11, 12	SRP118634	P	4-дневные проростки	10 мкМ АЦК, 2 ч; 10 мкМ АЦК, 4 ч	3	Fu et al., 2021
1, 2, 6, 7	SRA063695	P	3-дневные этилированные проростки	10 мл ⁻¹ этилен газ, 4 ч* 10 мл ⁻¹ этилен газ, 12 ч 10 мл ⁻¹ этилен газ, 24 ч	3	Chang et al., 2013
3	SRP069072	P	3-дневные этилированные проростки	10 мл ⁻¹ этилен газ, 4 ч	2	Zhang et al., 2016a
4, 8	SRP076862	P	Корни и побеги 3-дневных этилированных проростков отдельно	10 мл ⁻¹ этилен газ, 4 ч	2	Zhang et al., 2016b
5	SRP168223	P	3-дневные этилированные проростки	Этилен газ, 4 ч	2	Zander et al., 2019
9	GSE7432	M	Корни 3-дневных этилированных проростков	10 мл ⁻¹ этилен газ, 4 ч	2	Stepanova et al., 2007
10	SRP126162	P	Корни 6-дневных проростков	10 мл ⁻¹ этилен газ, 4 ч	2	Feng et al., 2017
13, 14, 15, 16	GSE84446	M	Корни 6-дневных проростков	1 мкМ АЦК, 4 ч; 1 мкМ АЦК, 8 ч; 1 мкМ АЦК, 12 ч; 1 мкМ АЦК, 24 ч	3	Harkey et al., 2018

* 1, 2 – эксперименты по обработке этиленом с одинаковыми условиями.

Количество биологических реплик, доступных для каждого образца и используемых для идентификации ДЭГ, указано в шестом столбце. Для обнаружения ДЭГ в экспериментах № 2, 4, 5, 7, 9, 16, 17, 22, 23 с ауксином и № 1, 2, 6, 7 с этиленом использовали необработанные контрольные образцы, собранные в начальный момент времени; в остальных случаях для каждого эксперимента были доступны отдельные ложно-обработанные контрольные образцы. P – РНК-секвенирование; M – микрочип эксперимент; 2,4-Д – 2,4-дихлорфеноксиуксусная кислота; НФК – нафтил-фталаминовая кислота; ИУК – индол-3-уксусная кислота; ИУК – 1-нафталинуксусная кислота; ИМК – индол-3-масляная кислота.

(Lawrence et al., 2013) и файл с аннотацией генома *A. thaliana*. Определение ДЭГ выполняли с помощью пакета DESeq2 v.1.34.0 (Love et al., 2014). Для каждого набора данных (как микрочипов, так и РНК-секвенирования) мы применили поправку на множественное тестирование гипотез по методу Бенджамини–Хохберга (Benjamini, Hochberg, 1995) для контроля доли ложноположительных результатов (false discovery rate, FDR) при определении дифференциально экспрессирующихся генов. Чтобы детектировать ДЭГ, мы использовали порог FDR, равный 0.05. В результате мы получили 23 списка ДЭГ после обработки ауксином и 16 – после обработки этиленом, в каждом из которых было не менее 300 ДЭГ (см. таблицу).

Результаты и обсуждение

В этой работе мы приводим пример использования InterTransViewer для сравнительного описания профилей дифференциальной экспрессии генов, отражающих транскрипционный ответ на фитогормоны у *A. thaliana*. Для анализа мы отобрали 23 профиля ауксин-индуцированной дифференциальной экспрессии генов из 16 различных исследований и 16 профилей дифференциальной экспрессии, индуцированной этиленом или его предшественником, 1-аминоциклопропановой кислотой (АЦК), из 8 исследований (см. таблицу и раздел «Материалы и методы»).

Рисунок 2 схематично иллюстрирует метаданные для каждого профиля дифференциально экспрессирующихся генов. Видно, что несмотря на общность исследуемого фактора, экспериментальные условия, в которых профили были получены, неоднородны. В частности, различались химическая природа активного вещества, его концентрация, способ и длительность обработки, условия выращивания растений, их возраст в момент сбора образцов, состав образцов, методы профилирования экспрессии. Только два ауксин-индуцированных профиля ДЭГ (№ 9 и 10) из двух исследований и три этилен-индуцированных профиля ДЭГ (№ 1, 2 и 3) также из двух исследований были получены в сходных условиях, согласно метаданным. Таким образом, целью дальнейшего сравнительного анализа стало исследование однородности фитогормон-индуцированных профилей ДЭГ в зависимости от условий, в которых они были получены.

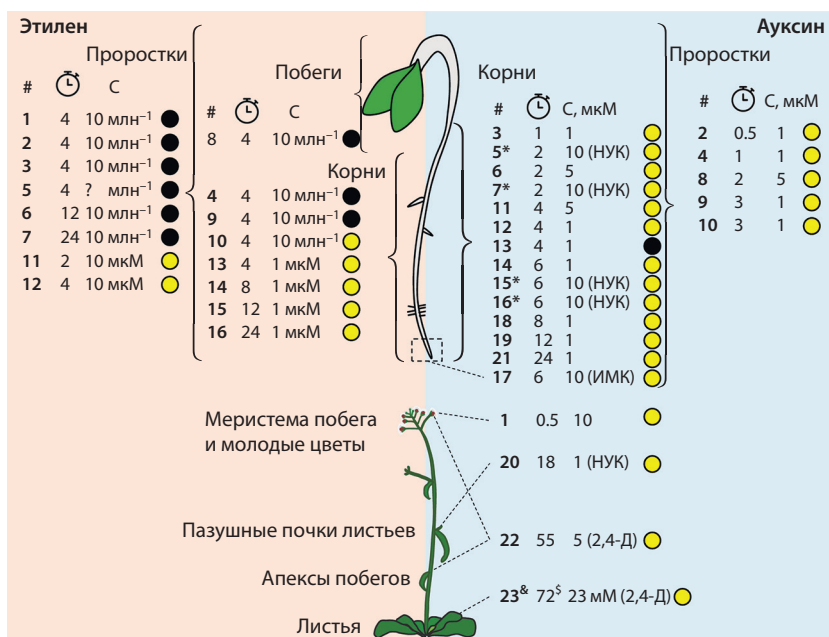


Рис. 2. Схематичное представление экспериментальных условий, в которых получены профили ДЭГ, выбранные для сравнительного анализа.

Звездочкой обозначен сегмент корня между апикальной меристемой корня и корневой шейкой (соединением корень-гипокотиль). Этилен применяли либо в виде газа (концентрация в частях на миллион), либо в виде его биологического предшественника АЦК (мкМ). Ауксин использован в форме ИУК, иначе указано. НУК – 1-нафталинуксусная кислота; ИМК – индолил-3-масляная кислота; 2,4-Д – 2,4-дихлорфеноксиуксусная кислота. Длительность обработки приведена в часах. Числа, выделенные жирным шрифтом, обозначают номера экспериментов из таблицы. С – концентрация гормона; черные кружки – этилированные проростки; желтые кружки – проростки, выращенные на свету; & – опрыскивание; § – после обработки; ? – концентрация газообразного этилена в первоисточнике не указана.

Ауксин- и этилен-индуцированные профили ДЭГ вариабельны по количеству ДЭГ

Прежде всего, мы охарактеризовали отдельные списки ДЭГ с помощью функции DEGsummary. Ауксин- и этилен-индуцированные профили ДЭГ оказались неоднородны по количеству ДЭГ: этот показатель варьировал от 410 до 11966 в ауксин-индуцированных профилях (медианное значение 3205) и от 379 до 5253 – в этилен-индуцированных профилях (медианное значение 1428) (рис. 3, а, б). Отклонение количества ДЭГ от значения медианы в большинстве случаев можно было объяснить специфическими условиями проведения экспериментов. Так, низкое количество ауксин-чувствительных ДЭГ наблюдалось в меристеме побега и молодых цветах после кратковременной обработки ауксином (№ 1; 586 ДЭГ), а также в корне при длительной обработке (24 ч) ауксином низкой концентрации (1 мкМ) в форме индол-3-уксусной кислоты (ИУК) (№ 21; 686 ДЭГ). Низкое количество ДЭГ в последнем случае связано с тем, что при ответе на ауксин пик изменений транскрипционной активности наблюдается при длительности обработки от 2 до 8 ч, в то время как при увеличении времени обработки до 12–24 ч транскрипционная активность большинства генов возвращается к уровню, наблюдаемому в контрольных (без обработки ауксином) образцах, а число генов с измененным уровнем транскрипции становится близким к числу генов при кратковременных (1 ч) обработках ауксином (Lewis et al., 2013). Высокое количество ДЭГ (№ 22, 11966 ДЭГ) было характерно для продолжительной обработки в течение 55 ч 5 мкМ 2,4-Д кончиков побегов и пазушных меристем с целью инициации каллусообразования, которое сопровождается значительным перепрограммированием транскрипционной активности генома (Xu et al., 2012).

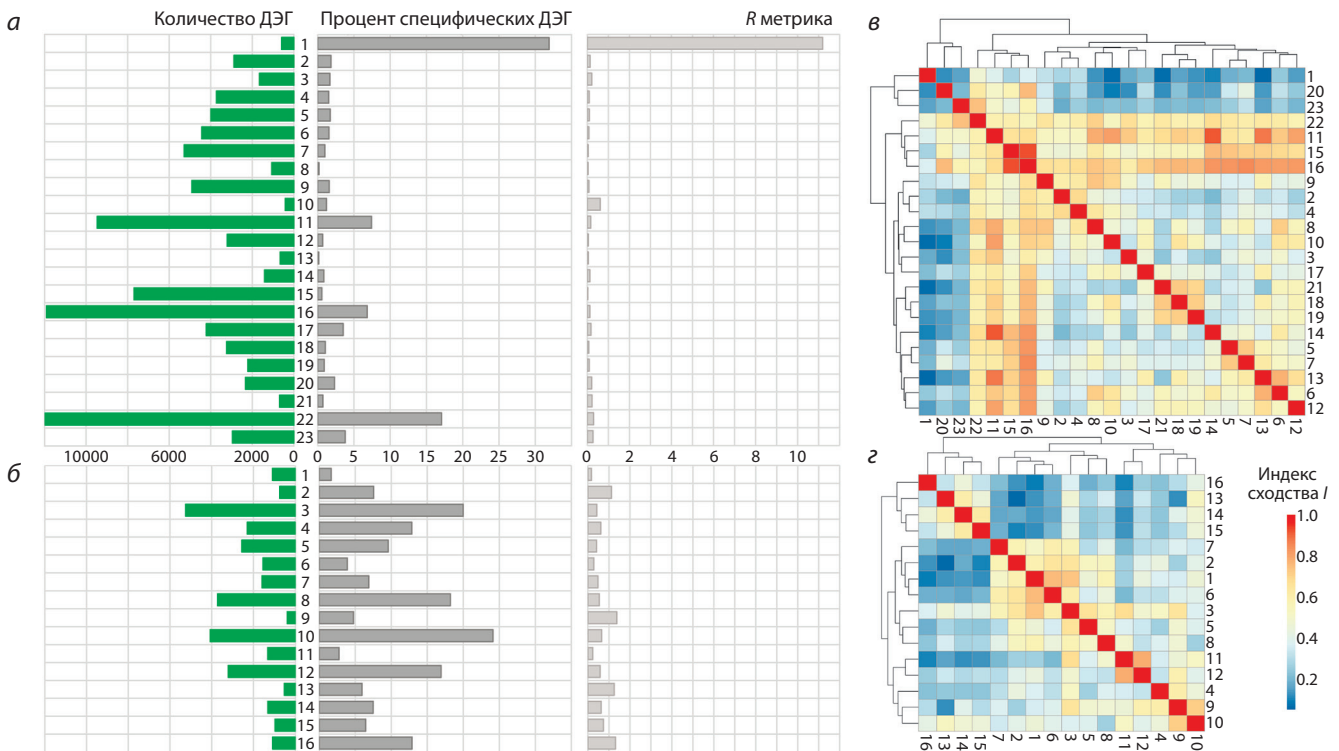


Рис. 3. Сравнительное описание транскрипционного ответа на ауксин и этилен в различных условиях у *A. thaliana*.

a, б – количество ДЭГ и доля специфических (уникальных) ДЭГ в наборах данных по ауксину (*a*) и этилену (*б*), а также *R* метрика для каждого набора данных; *в, з* – попарное сравнение экспериментов с ауксином (*в*) и экспериментов с этиленом (*з*). Индекс сходства *I*, отражающий долю общих ДЭГ в меньшем списке, описан в разделе «Материалы и методы». Номера экспериментов соответствуют номерам в таблице.

Достаточно большое количество ДЭГ наблюдалось и при менее продолжительной обработке проростков 5–10 мкМ ИУК в течение 4–6 ч, что соответствует пику изменений транскрипционной активности при ответе на ауксин (Lewis et al., 2013). Необходимо отметить, что большое количество ДЭГ наблюдалось в транскриптомах целых корней или корней без кончика корня с большим разнообразием тканей (№ 11, 15 и 16; 9461, 7692 и 11905 ДЭГ соответственно). Если исследовали, наоборот, только кончик корня (№ 17), то число ДЭГ уменьшалось до 4214, что можно объяснить достаточной биологической однородностью материала (колумелла, ниша ствольных клеток и первые потомки инициалей).

Следует отметить высокое значение отношения *R* для профиля ДЭГ меристемы побега и молодых цветов (№ 1), что свидетельствует о специфическом по составу ДЭГ транскрипционном ответе на ауксин в этих органах по сравнению со всеми остальными представленными в исследовании. Причиной существенного отклонения количества ДЭГ от медианы в профиле № 10 (410 ДЭГ) мог быть стресс из-за резкой смены условий выращивания проростков, когда перед обработкой ауксином проростки, выращенные в течение 6–7 дней на агаризованной среде, помещали на сутки в жидкую среду с постоянным перемешиванием на шейкере. В такой ситуации гены ауксинового ответа, ассоциированные с ответом на стресс, изменили свою экспрессию и в опыте, и в контроле, в результате после обработки ауксином остались ДЭГами только гены, не связанные со стрессом. В то же время, учитывая несколько

повышенное значение отношения *R* для профиля № 10 по сравнению с медианой, можно сделать предположение о невысоком качестве этих данных.

Низкое количество этилен-чувствительных ДЭГ наблюдалось в корнях выращенных на свету проростков после непродолжительной (4 ч) обработки биологическим предшественником этилена, АЦК, в низкой (1 мкМ) концентрации (№ 13; 522 ДЭГ), что может быть связано с малым временем обработки для реализации полного ответа на этилен. Увеличение времени обработки до 8, 12 и 24 ч (№ 14, 15, 16) во всех случаях увеличивало число ДЭГ примерно в 2 раза (Harkey et al., 2018).

Таким образом, начиная примерно с 8 ч при такой обработке реализуется полный ответ на этилен, при котором количество ДЭГ близко к медианному. Низкое количество ДЭГ в профиле № 9 (379 ДЭГ) можно связать с техническими особенностями транскриптомных экспериментов в работе (Stepanova et al., 2007), учитывая низкое количество ДЭГ в ауксин-индуцированном профиле № 13 (657 ДЭГ) из этого же исследования. Тем не менее нет оснований говорить о низком качестве этих данных, так как наблюдаемые отклонения не сопровождаются существенным повышением значения отношения *R*. Примечательно, что значения числа ДЭГ, близкие к медианному, дали РНК-секвенирование на SOLiD (№ 1, 2, 6, 7) вне зависимости от времени обработки (Chang et al., 2013), а также секвенирование с помощью Illumina побегов (№ 4) и растений другого экотипа (*Ler*, *Landsberg erecta*) (№ 5), а не экотипа *Columbia*, как во всех остальных случаях. При этом

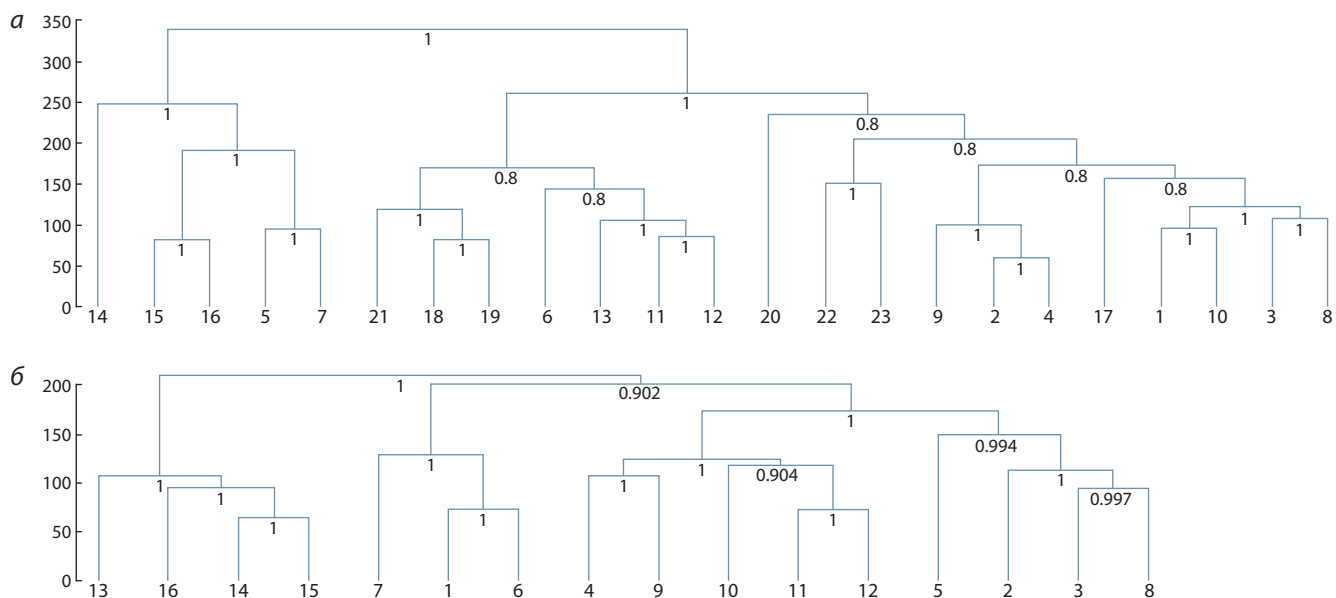


Рис. 4. Иерархическая кластеризация профилей дифференциальной экспрессии в ответ на ауксин (а) и этилен (б) с использованием метода Ward.D2.

Шумина секвенирование этиолированных проростков и корней дало число ДЭГ, значительно выше медианных (№ 3, 8 и 10; 5253, 3715 и 4067 ДЭГ соответственно).

Профили дифференциальной экспрессии генов в ответ на фитогормоны в образцах, представленных разными частями растения, отличаются по составу ДЭГ

Далее мы более подробно исследовали сходство профилей экспрессии по составу дифференциально экспрессирующихся генов. Парные сравнения с использованием функций GetSimMatrix подтвердили специфический характер транскрипционного ответа на ауксин в меристеме побега и молодых цветах (профиль № 1) по сравнению со всеми остальными органами (см. рис. 3, в). Неудивительно, что относительно высокое значение индекса сходства для этого профиля ($I = 0.47$) наблюдалось только в паре с профилем дифференциальной экспрессии генов в меристемах побегов (№ 22). Затем выделялись две группы сходных профилей, описывающих транскрипционный ответ на ауксин в целых проростках и в корнях. Разделение профилей, соответствующих целым проросткам и их корням, подтверждается также результатами кластеризации профилей по дифференциальной экспрессии с помощью функции DE_clustering и является очевидным, так как в проростках наряду с корнем присутствует также побег (рис. 4). Однако, несмотря на детектируемое внутригрупповое сходство, нужно отметить очевидную вариабельность профилей внутри каждой из этих двух групп (см. рис. 2, в). Наконец, профили с большим количеством ДЭГ (№ 11, 15, 16, 22) демонстрировали достаточно высокий индекс сходства при попарном сравнении со всеми профилями (см. рис. 3, в).

Качественное сходство профилей с большим количеством ДЭГ друг с другом, а также с более узкими списками ДЭГ свидетельствует об их корректности. Профили дифференциальной экспрессии генов на почках стеблевых

листьев (№ 20) и в листьях (№ 23) демонстрировали умеренное сходство ($I \geq 0.42$) только с профилями с большим количеством ДЭГ. Можно предположить, что обработка высокими концентрациями ауксина (№ 11, 15, 16, 22) изменяет экспрессию разных групп генов, каждая из которых отвечает на низкие концентрации ауксина только в определенных условиях. Помимо этого, большее число ДЭГ в позднем ответе может объясняться широкой представленностью в них генов вторичного ответа на ауксин.

Попарные сравнения этилен-индуцированных профилей дифференциальной экспрессии выявили обособленную группу (№ 13–16) из исследования А.Ф. Harkey с коллегами (2018) (см. рис. 2, з). Они описывали изменение экспрессии генов в корнях обработанных АЦК (предшественником этилена) проростков, растущих в условиях постоянной освещенности. Считается, что АЦК также имеет независимую от этилена биологическую активность (Vanderstraeten et al., 2019), а свет оказывает существенное влияние на формирование транскрипционного ответа на этилен у *A. thaliana* (Shi et al., 2016a, b; Luo, Shi, 2019). Мы предположили, что химическая природа активного вещества и условия освещенности при выращивании проростков могли выступать весомыми нецелевыми факторами в данном случае. Однако профили № 13–16 демонстрировали лишь умеренное сходство с профилем дифференциальной экспрессии в корнях обработанных этиленом проростков, выращенных при долгом (16 ч) дне (№ 10) ($I = 0.59, 0.54, 0.52$ и 0.49), и достаточно сильно отличались от профилей дифференциальной экспрессии генов в целых проростках, выращенных при 12-часовом дне и обработанных АЦК (№ 11 и 12) ($0.22 < I < 0.42$) (см. рис. 3, з). Таким образом, нельзя исключить, что обособленность профилей № 13–16 может быть следствием пакетного эффекта.

Остальные профили распадались на две группы сходных списков. Первые характеризовали транскрипцион-

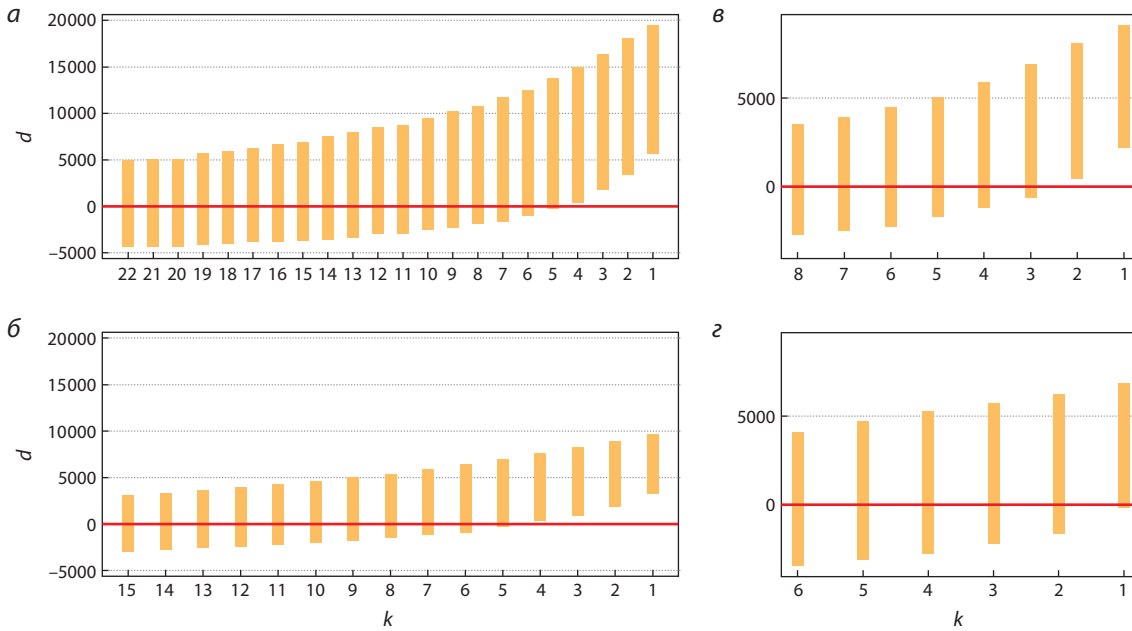


Рис. 5. Оценка однородности профилей дифференциальной экспрессии ответа на ауксин (а, б) и этилен (в, г) посредством ресемплинга.

Столбцы отражают 95 % доверительный интервал значений $d = N_m - N_k$, где N_m – число генов, которые являются ДЭГ хотя бы в одном из списков в псевдовыборке размера m ; N_k – число генов, которые определяются как ДЭГ хотя бы в одном из списков в псевдовыборке размера k , $m > k$. Детальное описание процедуры представлено в разделе «Материалы и методы». а – результаты расчетов для $m = 23$ (все ауксин-индуцированные профили ДЭГ); б – результаты расчетов для $m = 9$ (ауксин-индуцированные профили ДЭГ № 5–7, 12–14, 18, 19, 21); в – результаты расчетов для $m = 16$ (все этилен-индуцированные профили ДЭГ); г – результаты расчетов для $m = 7$ (ауксин-индуцированные профили ДЭГ № 1–3, 5–8). Жирной линией отмечено значение $d = 0$.

ный ответ на этилен в корнях проростков независимо от режима освещенности, а также в целых проростках, выращенных в присутствии света. Вторые описывали ответ на этилен в этилированных проростках или побегах. Таким образом, мы подтвердили известные факты, что свет играет существенную роль в формировании ответа на этилен (Shi et al., 2016a, b; Luo, Shi, 2019), но дополнительно показали, что этот эффект наблюдается в побегах, но не в корне. Следует отметить, что при кластеризации с учетом величины изменения экспрессии временной ряд по обработке этиленом этилированных проростков из исследования K.N. Chang с коллегами (2013) выделяется в отдельную группу (см. рис. 4), что также поднимает вопрос о возможном пакетном эффекте.

Набор из семи этилен-индуцированных профилей является однородным по составу ДЭГ

Количество генов, идентифицируемых как ДЭГ хотя бы в одном из анализируемых списков, в существенной мере определяется однородностью набора профилей. В нашем случае 20552 и 10988 генов были определены как ДЭГ по крайней мере в одном профиле дифференциальной экспрессии, индуцированной ауксином и этиленом/АЦК соответственно. С учетом размера генома *A. thaliana*, который содержит немногим более 30000 генов, это неожиданно большое количество ДЭГ, которое заметно превышает количество ДЭГ в отдельных экспериментах и, вероятно, обусловлено изменчивостью индукции транскриптомов в зависимости от условий проведения экспериментов. Количественная оценка однородности наборов профилей

посредством ресемплинга (с использованием функции DE_bootstrap) ожидаемо показала их неоднородность по составу ДЭГ (рис. 5, а, б).

В то же время на основании описанных в предыдущем разделе результатов попарного сравнения профилей (см. рис. 3, в, г) можно сделать предположение о потенциальной однородности ауксин-индуцированных профилей ДЭГ в корне (№ 5–7, 12–14, 18–21) и этилен-индуцированных профилей ДЭГ в проростках/побегах (№ 1–3, 5–8). Чтобы проверить это предположение, мы осуществили анализ соответствующих наборов профилей дифференциальной экспрессии с применением функции DE_bootstrap. В то время как группа профилей ответа на ауксин в корнях все еще демонстрировала неоднородность (вероятно, длительность обработки обуславливала различия по составу ДЭГ), значимых отличий количества этилен-индуцированных ДЭГ в этилированных проростках не обнаруживалось. Таким образом, набор этилен-индуцированных профилей ДЭГ в этилированных проростках/побегах (№ 1–3, 5–8) в силу их однородности целесообразно в дальнейшем использовать для метаанализа (например, с целью более эффективного выявления слабых закономерностей).

Заключение

Метаанализ транскриптомных данных открывает широкие возможности для увеличения мощности статистического анализа при условии их однородности. Однако осмысленный выбор экспериментов для метаанализа зачастую затруднен отсутствием стандартов в этой области и удобных программных средств для сравнительной характери-

стики профилей дифференциальной экспрессии генов, в особенности для построения интуитивно понятных визуализаций. В настоящей работе мы предложили набор количественных индикаторов для сравнительной характеристики профилей ДЭГ (n – количество ДЭГ; δ – доля ДЭГ, специфичных для данного списка; R – отношение, характеризующее специфичность транскрипционного ответа; I – индекс сходства пары списков по составу ДЭГ; оценка однородности профилей ДЭГ), а также реализовали их вычисление и визуализацию в виде программы InterTransViewer. Продемонстрировано, что комплексное рассмотрение характеристик отдельных профилей ДЭГ (n , δ , R) с учетом результатов попарных сравнений профилей по составу ДЭГ (как с использованием индекса сходства I , так и посредством кластеризации с учетом величины изменения уровней экспрессии) позволяет позиционировать эксперименты в контексте друг друга, оценивать тенденцию к их интеграции или сегрегации, генерировать гипотезы о влиянии весомых нецелевых факторов на исследуемый транскрипционный ответ. В результате это дает возможность выделять потенциально однородные группы экспериментов.

Последующий анализ однородности этих групп профилей с применением процедуры ресемплинга и установления порога уровня значимости позволяет принять решение о целесообразности использования этих данных для метаанализа. В целом программа InterTransViewer дает возможность эффективно формировать выборки экспериментов в зависимости от задачи и методов метаанализа.

Список литературы / References

- Bairakdar M.D., Tewari A., Truttmann M.C. A meta-analysis of RNA-Seq studies to identify novel genes that regulate aging. *Exp. Gerontol.* 2023;173:112107. DOI 10.1016/j.exger.2023.112107
- Bargmann B.O., Vanneste S., Krouk G., Nawy T., Efroni I., Shani E., Choe G., Friml J., Bergmann D.C., Estelle M., Birnbaum K.D. A map of cell type-specific auxin responses. *Mol. Syst. Biol.* 2013; 9:688. DOI 10.1038/msb.2013.40
- Benjamini Y., Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B. Methodol.* 1995;57(1):289-300. DOI 10.1111/j.2517-6161.1995.tb02031.x
- Brazma A. Minimum information about a microarray experiment (MIAME) – successes, failures, challenges. *Sci. World J.* 2009; 29(9):420-423. DOI 10.1100/tsw.2009.57
- Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C., Gaasterland T., Glenisson P., Holstege F.C., Kim I.F., Markowitz V., Matese J.C., Parkinson H., Robinson A., Sarkans U., Schulze-Kremer S., Stewart J., Taylor R., Vilo J., Vingron M. Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat. Genet.* 2001;29(4):365-371. DOI 10.1038/ng1201-365
- Cahan P., Rovegno F., Mooney D., Newman J.C., St. Laurent G. 3rd, McCaffrey T.A. Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization. *Gene.* 2007; 401(1-2):12-18. DOI 10.1016/j.gene.2007.06.016
- Chaiwanon J., Wang Z.Y. Spatiotemporal brassinosteroid signaling and antagonism with auxin pattern stem cell dynamics in *Arabidopsis* roots. *Curr. Biol.* 2015;25(8):1031-1042. DOI 10.1016/j.cub.2015.02.046
- Chang K.N., Zhong S., Weirauch M.T., Hon G., Pelizzola M., Li H., Huang S.S., Schmitz R.J., Urlich M.A., Kuo D., Nery J.R., Qiao H., Yang A., Jamali A., Chen H., Ideker T., Ren B., Bar-Joseph Z., Hughes T.R., Ecker J.R. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in *Arabidopsis*. *eLife.* 2013;2:e00675. DOI 10.7554/eLife.00675
- Chen S., Zhou Y., Chen Y., Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884-i890. DOI 10.1093/bioinformatics/bty560
- Clough E., Barrett T. The gene expression omnibus database. *Methods Mol. Biol.* 2016;1418:93-110. DOI 10.1007/978-1-4939-3578-9_5
- De Rybel B., Audenaert D., Xuan W., Overvoorde P., Strader L.C., Kepinski S., Hoye R., Brisbois R., Parizot B., Vanneste S., Liu X., Gilday A., Graham I.A., Nguyen L., Jansen L., Njo M.F., Inzé D., Bartel B., Beeckman T. A role for the root cap in root branching revealed by the non-auxin probe naxillin. *Nat. Chem. Biol.* 2012;8(9): 798-805. DOI 10.1038/nchembio.1044
- Delker C., Pöschl Y., Raschke A., Ullrich K., Ettingshausen S., Hauptmann V., Grosse I., Quint M. Natural variation of transcriptional auxin response networks in *Arabidopsis thaliana*. *Plant Cell.* 2010; 22(7):2184-2200. DOI 10.1105/tpc.110.073957
- Feng Y., Xu P., Li B., Li P., Wen X., An F., Gong Y., Xin Y., Zhu Z., Wang Y., Guo H. Ethylene promotes root hair growth through coordinated EIN3/EIL1 and RHD6/RSL1 activity in *Arabidopsis*. *Proc. Natl. Acad. Sci.* 2017;114:13834-13839. DOI 10.1073/pnas.1711723115
- Freire-Rios A., Tanaka K., Crespo I., van der Wijk E., Sizentsova Ya., Levitsky V., Lindhoud S., Fontana M., Hohlbein J., Boer R., Mironova V., Weijers D. Architecture of DNA elements mediating ARF transcription factor binding and auxin-responsive gene expression in *Arabidopsis*. *Proc. Natl. Acad. Sci.* 2020;117(39):24557-24566. DOI 10.1073/pnas.2009554117
- Fu L., Liu Y., Qin G., Wu P., Zi H., Xu Z., Zhao X., Wang Y., Li Y., Yang S., Peng C., Wong C.C.L., Yoo S.D., Zuo Z., Liu R., Cho Y.H., Xiong Y. The TOR–EIN2 axis mediates nuclear signalling to modulate plant growth. *Nature.* 2021;591(7849):288-292. DOI 10.1038/s41586-021-03310-y
- Harkey A.F., Watkins J.M., Olex A.L., DiNapoli K.T., Lewis D.R., Fetrow J.S., Binder B.M., Muday G.K. Identification of transcriptional and receptor networks that control root responses to ethylene. *Plant Physiol.* 2018;176(3):2095-2118. DOI 10.1104/pp.17.00907
- Kang D.D., Sibille E., Kaminski N., Tseng G.C. MetaQC: objective quality control and inclusion/exclusion criteria for genomic meta-analysis. *Nucleic Acids Res.* 2012;40(2):e15. DOI 10.1093/nar/gkr1071
- Keel B.N., Lindholm-Perry A.K. Recent developments and future directions in meta-analysis of differential gene expression in livestock RNA-Seq. *Front. Genet.* 2022;19(13):983043. DOI 10.3389/fgene.2022.983043
- Kim D., Pertege G., Trapnell C., Pimentel H., Kelley R., Salzberg S.L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14:R36. DOI 10.1186/gb-2013-14-4-r36
- Kim D., Paggi J.M., Park C., Bennett C., Salzberg S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 2019;37:907-915. DOI 10.1038/s41587-019-0201-4
- Lawrence M., Huber W., Pagès H., Aboyoun P., Carlson M., Gentleman R., Morgan M.T., Carey V.J. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 2013;9:e1003118. DOI 10.1371/journal.pcbi.1003118
- Lewis D.R., Olex A.L., Lundy S.R., Turkett W.H., Fetrow J.S., Muday G.K. A kinetic analysis of the auxin transcriptome reveals cell wall remodeling proteins that modulate lateral root development in *Arabidopsis*. *Plant Cell.* 2013;25:3329-3346. DOI 10.1105/tpc.113.114868
- Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15(12):550. DOI 10.1186/s13059-014-0550-8
- Luo Y., Shi H. Direct regulation of phytohormone actions by photoreceptors. *Trends Plant Sci.* 2019;24:105-108. DOI 10.1016/j.tplants.2018.11.002

- McDermaid A., Monier B., Zhao J., Liu B., Ma Q. Interpretation of differential gene expression results of RNA-seq data: review and integration. *Brief. Bioinform.* 2019;20(6):2044-2054. DOI 10.1093/bib/bby067
- Mozgová I., Muñoz-Viana R., Hennig L. PRC2 represses hormone-induced somatic embryogenesis in vegetative tissue of *Arabidopsis thaliana*. *PLoS Genet.* 2017;13(1):e1006562. DOI 10.1371/journal.pgen.1006562
- Müller D., Waldie T., Miyawaki K., To J.P., Melnyk C.W., Kieber J.J., Kakimoto T., Leyser O. Cytokinin is required for escape but not release from auxin mediated apical dominance. *Plant J.* 2015;82(5): 874-886. DOI 10.1111/tpj.12862
- Okushima Y., Overvoorde P.J., Arima K., Alonso J.M., Chan A., Chang C., Ecker J.R., Hughes B., Lui A., Nguyen D., Onodera C., Quach H., Smith A., Yu G., Theologis A. Functional genomic analysis of the *AUXIN RESPONSE FACTOR* gene family members in *Arabidopsis thaliana*: unique and overlapping functions of *ARF7* and *ARF19*. *Plant Cell.* 2005;17(2):444-463. DOI 10.1105/tpc.104.028316
- Omelyanchuk N.A., Wiebe D.S., Novikova D.D., Levitsky V.G., Klimova N., Gorelova V., Weinholdt C., Vasiliev G.V., Zemlyanskaya E.V., Kolchanov N.A., Kochetov A.V. Auxin regulates functional gene groups in a fold-change-specific manner in *Arabidopsis thaliana* roots. *Sci. Rep.* 2017;7(1):2489. DOI 10.1038/s41598-017-02476-8
- Ritchie M.E., Phipson B., Wu D., Hu Y., Law C.W., Shi W., Smyth G.K. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47. DOI 10.1093/nar/gkv007
- Romero-Puertas M.C., Peláez-Vico M.Á., Pazmiño D.M., Rodríguez-Serrano M., Terrón-Camero L., Bautista R., Gómez-Cadenas A., Claros M.G., León J., Sandalio L.M. Insights into ROS-dependent signalling underlying transcriptomic plant responses to the herbicide 2,4-D. *Plant Cell Environ.* 2022;45(2):572-590. DOI 10.1111/pce.14229
- Rousseeuw P.J., Pernet C.R., Wilcox R.R. The percentile bootstrap: a primer with step-by-step instructions in R. *Adv. Meth. Pract. Psychol. Sci.* 2021;4(1). DOI 10.1177/2515245920911881
- Rung J., Brazma A. Reuse of public genome-wide gene expression data. *Nat. Rev. Genet.* 2013;14(2):89-99. DOI 10.1038/nrg3394
- Sarkans U., Füllgrabe A., Ali A., Athar A., Behrangi E., Diaz N., Fexova S., George N., Iqbal H., Kurri S., Munoz J., Rada J., Papatheodorou I., Brazma A. From arrayexpress to biostudies. *Nucleic Acids Res.* 2021;49(D1):D1502-D1506. DOI 10.1093/nar/gkaa1062
- Shi H., Liu R., Xue C., Shen X., Wei N., Deng X.W., Zhong S. Seedlings transduce the depth and mechanical pressure of covering soil using COP1 and ethylene to regulate EBF1/EBF2 for soil emergence. *Curr. Biol.* 2016a;26(2):139-149. DOI 10.1016/j.cub.2015.11.053
- Shi H., Shen X., Liu R., Xue C., Wei N., Deng X.W., Zhong S. The red light receptor phytochrome B directly enhances substrate-E3 ligase interactions to attenuate ethylene responses. *Dev. Cell.* 2016b;39(5): 597-610. DOI 10.1016/j.devcel.2016.10.020
- Simonini S., Bencivenga S., Trick M., Østergaard L. Auxin-induced modulation of ETTIN activity orchestrates gene expression in *Arabidopsis*. *Plant Cell.* 2017;29(8):1864-1882. DOI 10.1105/tpc.17.00389
- Stelpflug S.C., Sekhon R.S., Vaillancourt B., Hirsch C.N., Buell C.R., de Leon N., Kaeppler S.M. An expanded maize gene expression atlas based on RNA sequencing and its use to explore root develop-
ment. *Plant Genome.* 2016;9(1):1-16. DOI 10.3835/plantgenome2015.04.0025
- Stepanova A.N., Yun J., Likhacheva A.V., Alonso J.M. Multilevel interactions between ethylene and auxin in *Arabidopsis* roots. *Plant Cell.* 2007;19(7):2169-2185. DOI 10.1105/tpc.107.052068
- Sudmant P.H., Alexis M.S., Burge C.B. Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biol.* 2015; 16:287. DOI 10.1186/s13059-015-0853-4
- Tello-Ruiz M.K., Stein J., Wei S., Preece J., Olson A., Naithani S., Amarasinghe V., Dharmawardhana P., Jiao Y., Mulvaney J., Kumari S., Chougule K., Elser J., Wang B., Thomason J., Bolser D.M., Kerhornou A., Walts B., Fonseca N.A., Huerta L., Keays M., Tang Y.A., Parkinson H., Fabregat A., McKay S., Weiser J., D'Eustachio P., Stein L., Petryszak R., Kersey P.J., Jaiswal P., Ware D. Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.* 2016;44(D1):D1133-D1140. DOI 10.1093/nar/gkv1179
- Vanderstraeten L., Depaepe T., Bertrand S., Van Der Straeten D. The ethylene precursor ACC affects early vegetative development independently of ethylene signaling. *Front. Plant Sci.* 2019;10:1591. DOI 10.3389/fpls.2019.01591
- Vanneste S., De Rybel B., Beemster G.T., Ljung K., De Smet I., Van Isterdael G., Naudts M., Iida R., Gruissem W., Tasaka M., Inzé D., Fukaki H., Beeckman T. Cell cycle progression in the pericycle is not sufficient for SOLITARY ROOT/IAA₁₄-mediated lateral root initiation in *Arabidopsis thaliana*. *Plant Cell.* 2005;17(11):3035-3050. DOI 10.1105/tpc.105.035493
- Winter C., Kosch R., Ludlow M., Osterhaus A.D.M.E., Jung K. Network meta-analysis correlates with analysis of merged independent transcriptome expression data. *BMC Bioinformatics.* 2019;20(1): 144. DOI 10.1186/s12859-019-2705-9
- Xu K., Liu J., Fan M., Xin W., Hu Y., Xu C. A genome-wide transcriptome profiling reveals the early molecular events during callus initiation in *Arabidopsis* multiple organs. *Genomics.* 2012;100(2):116-124. DOI 10.1016/j.ygeno.2012.05.013
- Xuan W., Audenaert D., Parizot B., Möller B.K., Njo M.F., De Rybel B., De Rop G., Van Isterdael G., Mähönen A.P., Vanneste S., Beeckman T. Root cap-derived auxin pre-patterns the longitudinal axis of the *Arabidopsis* root. *Curr. Biol.* 2015;25(10):1381-1388. DOI 10.1016/j.cub.2015.03.046
- Yu X.T., Zeng T. Integrative analysis of omics big data. *Methods Mol. Biol.* 2018;1754:109-135. DOI 10.1007/978-1-4939-7717-8_7
- Zander M., Willige B.C., He Y., Nguyen T.A., Langford A.E., Nehring R., Howell E., McGrath R., Bartlett A., Castanon R., Nery J.R., Chen H., Zhang Z., Jupe F., Stepanova A., Schmitz R.J., Lewsey M.G., Chory J., Ecker J.R. Epigenetic silencing of a multifunctional plant stress regulator. *eLife.* 2019;8:e47835. DOI 10.7554/eLife.47835
- Zhang F., Qi B., Wang L., Zhao B., Rode S., Riggan N.D., Ecker J.R., Qiao H. EIN2-dependent regulation of acetylation of histone H3K14 and non-canonical histone H3K23 in ethylene signalling. *Nat. Commun.* 2016a;7(1):13018. DOI 10.1038/ncomms13018
- Zhang F., Wang L., Lim J.Y., Kim T., Pyo Y., Sung S., Shin C., Qiao H. Phosphorylation of CBP20 links microRNA to root growth in the ethylene response. *PLoS Genet.* 2016b;12(11):e1006437. DOI 10.1371/journal.pgen.1006437
- Zhou G., Soufan O., Ewald J., Hancock R.E.W., Basu N., Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* 2019;47(W1):W234-W241. DOI 10.1093/nar/gkz240

ORCID

A.V. Tyapkin orcid.org/0000-0002-5969-3628
V.V. Lavrekha orcid.org/0000-0001-8813-8941

D.Yu. Oshchepkov orcid.org/0000-0002-6097-5155
E.V. Zemlyanskaya orcid.org/0009-0005-7316-7690

Благодарности. Разработка метода выполнена в рамках бюджетного проекта FWNR-2022-0020. Сравнительный анализ транскриптомов выполнен при поддержке Российского научного фонда, грант № 20-14-00140.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 16.08.2023. После доработки 18.10.2023. Принята к публикации 02.11.2023.